CLAIMS

[1]   A voice synthesis device, comprising:

a memory unit operable to store, in advance, first voice
5   element information regarding plural voice elements having a first
voice quality, and second voice element information regarding plural
voice elements having a second voice quality that is different from
the first voice quality;

a voice information generating unit operable to acquire text
10   data, to generate, from the first voice element information in said
memory unit, first synthetic voice information indicating synthetic
voice having the first voice quality which corresponds to a character
that is included in the text data, and to generate, from the second
voice element information in said memory unit, second synthetic
15   voice information indicating synthetic voice having the second voice
quality which corresponds to a character that is included in the text
data;

a morphing unit operable to generate, from the first and
second synthetic voice information generated by said voice
20   information generating unit, intermediate synthetic voice
information indicating synthetic voice having intermediate voice
quality between the first and second voice quality which each
corresponds to a character that is included in the text data; and

a voice outputting unit operable to convert, to synthetic voice
25   having the intermediate voice quality, the intermediate synthetic
voice information generated by said morphing unit, and to output
the resulting synthetic voice,

wherein said voice information generating unit is operable to
generate each of the first and second synthetic voice information as
30   a sequence of plural characteristic parameters, and

said morphing unit is operable to generate the intermediate
synthetic voice information by calculating an intermediate value of

characteristic parameters to which the first and second synthetic voice information respectively correspond.

[2]   The voice synthesis device according to Claim 1,
wherein said morphing unit is operable to change the ratio of contribution of the first and second synthetic voice information to the intermediate synthetic voice information so that the voice quality of the synthetic voice outputted from said voice outputting unit continuously changes during the output of the synthetic voice.

[3]   The voice synthesis device according to Claim 1,
wherein said memory unit is operable to store characteristic information which indicates a standard in each voice element that is indicated by each of the first and second voice element information in such a manner that the characteristic information is included in each of the first and second voice element information,
said voice information generating unit is operable to generate the first and second synthetic voice information in such a manner that the characteristic information is included in each of the first and second synthetic voice information, and
said morphing unit is operable to match the first and second synthetic voice information using the standard that is indicated by the characteristic information which is included in each of the first and second synthetic voice information, and to generate the intermediate synthetic voice information.

[4]   The voice synthesis device according to Claim 3,
wherein the standard is a point at which an acoustic characteristic of each voice element that is indicated by each of the first and second voice element information changes.

[5]   The voice synthesis device according to Claim 4,

wherein the point at which the acoustic characteristic changes is a state transition point on the most likely path in which each of the voice element indicated by each of the first and second voice element information is represented by a hidden Markov model (HMM), and

said morphing unit is operable to match the first and second synthetic voice information along the time axis using the state transition point, and to generate the intermediate synthetic voice information.

[6]   The voice synthesis device according to Claim 1, further comprising:

an image storing unit operable to store, in advance, first image information indicating an image which corresponds to the first voice quality and second image information indicating an image which corresponds to the second voice quality;

an image morphing unit operable to generate, from the first and second image information, intermediate image information indicating an intermediate image of images which are respectively indicated by the first and second image information, the intermediate image information indicating an image which corresponds to the voice quality of the intermediate synthetic sound information; and

a display unit operable to acquire intermediate image information generated by said image morphing unit, and to display an image that is indicated by the intermediate image information in synchronization with synthetic voice outputted from said voice outputting unit.

[7]   The voice synthesis device according to Claim 6,

wherein the first image information indicates a face image which corresponds to the first voice quality and the second image

information indicates a face image which corresponds to the second voice quality.

[8]   The voice synthesis device according to Claim 1, further comprising:

a designating unit operable to place, at respective N^{th} dimensional coordinates for display where N is a natural number, fixed points indicating the first and second voice quality and moving points which move on the basis of operation by a user, to derive the ratio of contribution of the first and second synthetic voice information to the intermediate synthetic voice information on the basis of the arrangement of the fixed points and moving points, and to designate the derived ratio to said morphing unit, and

said morphing unit is operable to generate the intermediate synthetic voice information in accordance with the ratio designated by said designation unit.

[9]   The voice synthesis device according to Claim 1,

wherein said voice information generating unit is operable to sequentially generate each of the first and second synthetic voice information.

[10]   The voice synthesis device according to Claim 1,

wherein said voice information generating unit is operable to generate each of the first and second synthetic voice information in parallel.

[11]   A voice synthesis method for generating and outputting synthetic voice using a memory which stores first voice element information on plural voice elements having first voice quality and second voice element information on plural voice elements having second voice quality that is different from the first voice quality in

advance, said voice synthesis method comprising:

a text acquiring step of acquiring text data;

a voice information generating step of generating, from the first voice element information of the memory, first synthetic voice information which indicates synthetic voice of the first voice quality which corresponds to a character that is included in the text data, and generating, from the second voice element information of the memory, second synthetic voice information which indicates synthetic voice of the second voice quality which corresponds to a character that is included in the text data;

a morphing step of generating, from the first and second synthetic voice information generated in said voice information generating step, intermediate synthetic voice information which indicates synthetic voice having intermediate voice quality between the first and second voice quality which corresponds to a character that is included in the text data; and

a voice outputting step of converting the intermediate synthetic voice information generated in said morphing step to synthetic voice having the intermediate voice quality and outputting the resulting synthetic voice,

wherein each of the first and second synthetic voice information is generated as a sequence of plural characteristic parameters in said voice information generating step, and

the intermediate synthetic voice information is generated by calculating an intermediate value of characteristic parameters which respectively correspond to the first and second synthetic voice information in said morphing step.

[12] The voice synthesis method according to Claim 11,

wherein the ratio of contribution of the first and second synthetic voice information to the intermediate synthetic voice information is changed in said morphing step, so that the voice

quality of the synthetic voice outputted in said voice outputting step continuously changes during output of the synthetic voice.

[13]   The voice synthesis method according to Claim 11,
wherein the memory stores characteristic information which indicates the standard in each voice element which is indicated by each of the first and second voice element information, in such a manner that the characteristic information is included in each of the first and second voice element information,

the first and second synthetic voice information is generated in such a manner that the characteristic information is included in each of the first and second synthetic voice information in said voice information generating step, and

the first and second synthetic voice information is matched using the standard that is indicated by the characteristic information included in each of the first and second synthetic voice information, and after that, the intermediate synthetic voice information is generated in said morphing step.

[14]   The voice synthesis method according to Claim 13,
wherein the standard is a point at which the acoustic characteristic of each voice element that is indicated by each of the first and second voice element information changes.

[15]   The voice synthesis method according to Claim 14,
wherein the point at which the acoustic characteristic changes is a point at which the state transits along the most likely course where each voice element that is indicated by each of the first and second voice element information is indicated by HMM (hidden Markov model), and

the first and second synthetic voice information is matched along the time axis using the point at which the state transits, and

after that, the intermediate synthetic voice information is generated in said morphing step.

[16]   The voice synthesis method according to Claim 11, further comprising:

an image morphing step of generating, from the first and second image information of an image memory which stores, in advance, first image information indicating an image which corresponds to the first voice quality and second image information indicating an image which corresponds to the second voice quality, intermediate image information indicating an intermediate image between images which are respectively indicated by the first and second image information by using the image memory, the intermediate image information indicating an image which corresponds to the voice quality of the intermediate synthetic sound information; and

a displaying step of displaying the image which is generated in said image morphing step and indicated by the intermediate image information in sync with synthetic voice outputted in said voice outputting step.

[17]   The voice synthesis method according to Claim 16,

wherein the first image information indicates a face image which corresponds to the first voice quality and the second image information indicates a face image which corresponds to the second voice quality.

[18]   A program for generating and outputting synthetic voice using a memory which stores first voice element information on plural voice elements having first voice quality and second voice element information on plural voice elements having second voice quality that is different from the first voice quality in advance, said

program causing a computer to execute:

a text acquiring step of acquiring text data;

a voice information generating step of generating, from the first voice element information of the memory, first synthetic voice information which indicates synthetic voice of the first voice quality which corresponds to a character that is included in the text data, and generating, from the second voice element information of the memory, second synthetic voice information which indicates synthetic voice of the second voice quality which corresponds to a character that is included in the text data;

a morphing step of generating, from the first and second synthetic voice information generated in the voice information generating step, intermediate synthetic voice information which indicates synthetic voice having intermediate voice quality between the first and second voice quality which corresponds to a character that is included in the text data; and

a voice outputting step of converting the intermediate synthetic voice information generated in the morphing step to synthetic voice having the intermediate voice quality and outputting the resulting synthetic voice, and

each of the first and second synthetic voice information is generated as a sequence of plural characteristic parameters in the voice information generating step, and

the intermediate synthetic voice information is generated by calculating an intermediate value of characteristic parameters which respectively correspond to the first and second synthetic voice information in the morphing step.